DOCUMENT RESUME

ED 147 306 80 SP 011 943

AUTHOR Andrews, Theodore E., Ed.
TITLE PBTE (Performance-Based Teacher Education); Vol, 2, No. 9, March 1974.
INSTITUTION Multi-State Consortium on Performance-Based Teacher Education, Albany, N.Y.
SPONS AGENCY Office of Education (DHEW), Washington, D.C.
PUB DATE Mar 74
NOTE 13p.

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *Academic Achievement; Learning Modules; *Literature Reviews; *Performance Based Teacher Education; *Student Ability; *Success Factors; Teacher Attitudes; Teacher Behavior; *Teacher Influence; Training Techniques
IDENTIFIERS California Beginning Teacher Evaluation Study; Florida Modules on Generic Teaching Competencies

ABSTRACT

This issue of the Multi-State Consortium on Performance-Based Teacher Education (PBTE) newsletter focuses on the role that assessment, research, and evaluation play in PBTE. David A. Potter presents a detailed examination of the literature investigating teacher behavior and student achievement, providing summaries of findings and recommendations for further research efforts. Marlaine Lockheed-Katz presents a review of studies relating teacher expectancy to teacher behavior and/or student performance. An overview is given of the California Beginning Teacher Evaluation Study, the first study in the United States attempting to systematically determine teacher behaviors affecting student achievement, with results intended for use in establishing state licensing procedures. A report is made on the experimental work performed in evaluating thirty training modules for inservice teachers developed by the Middle School Teaching Competencies project at the University of Florida in the Florida Modules on Generic Teaching Competencies service. (MJB)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

PB 0989

ED147306

PBTE

PERFORMANCE - BASED TEACHER EDUCATION

Published by the Multi-State Consortium on Performance-Based Teacher Education

Vol. 2, No. 9 March 1974

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Assessment, Research, Evaluation

PBTE is pleased to be able to bring to its readers this special issue on competency assessment, research, and evaluation. These themes are also the topics for a major conference this month at the University of Houston.

Possibly the one unifying concern shared by everyone in performance-based (competence-based) teacher education is assessment. How does one measure teacher effectiveness? Critics of performance education say that research must be done to establish relationships between teacher competencies and student learning. We agree.

This issue of PBTE attempts to show where we are. We hope that readers will share with us the excitement of finding that some answers are beginning to emerge, and will conclude that this major undertaking has, at least, begun.

The accompanying articles by David Potter and Marlaine Lockheed-Katz were made available to PBTE by Educational Testing Service. They appear in "A Critical Review of the Literature: Teacher Performance and Pupil Growth," December 1973, which contains a series of papers prepared for the California Commission on Teacher Preparation and Licensing's "Beginning Teacher Evaluation Study." PBTE is grateful to Frederick J. McDonald, project director, for permission to reprint these articles.

Teacher Behavior and Student Achievement

by David A. Potter

In a world full of educators and educational researchers, there is an amazing dearth of empirically derived knowledge concerning the relationship between teacher behavior and student achievement. Rosenshine and Furst (1971) began their review of the literature in this area with the statement, "This review is an admission that we know very little about the relationship between classroom behavior and student gains. It is a plea for more research . . ." Their statement was a general one, as was their review; it spanned many grade levels, content areas, and different populations. A reviewer who limits himself to the teaching of reading and mathematics at the elementary level is hardly likely to reach conclusions more optimistic than those of Rosenshine and Furst; if little is known in general, still less is known about the relationship between teacher behavior and student achievement in reading and mathematics at the elementary level.

Our review generally supports this pessimism. The state of the art is illustrated by the studies cited in three recent reviews. Weintraub, Robinson, Smith, and Plessas (1973) reviewed 302 reports of investigations relating to reading during the period from July 1, 1971 to June 30, 1972. Of these, 32 dealt with the teaching of reading to elementary school students. Not one single study reports or investigates the relationship between teacher behavior and student achievement. Instead, the majority focused on the differential effectiveness of various programs for the teaching of reading (for example, i.t.a. vs. traditional orthography) or on educational technology (e.g., programmed instruction). Only rarely do such studies even attempt to determine the ways in which participating teachers implemented the programs, leaving the inquiring reader to wonder how much of the variance in student achievement across programs could have been accounted for by variance in teacher behavior.

The recent *Second Handbook of Research on Teaching* (Travers, 1973) throws little additional light into these dark corners. Della-Piana and Endo (1973) provide a review of reading research; their section headings tell the story — comparative treatment studies of beginning reading programs, reading processes, and the technology of reading instruction. Only one implication for teacher behavior emerges; there is a definite tendency across studies for meaning emphasis methods (e.g., the Basal Approach) to lead to lower levels of student achievement than code emphasis methods (e.g., phonic methods). But the results are by no means conclusive, and here again there is a universal lack of attention to teacher behavior as an important variable in research on instruction.

(Continued on page 2)

47771

SP011 943

Teacher . . . *(continued)*

A review of research on the teaching of elementary school mathematics (Riedesel and Burns, 1973) reflects a similar emphasis. Again, programmatic and technological studies predominate.

But if the research literature is weak in terms of definitive answers in this area, it nevertheless does indicate that certain research directions seem more likely to be fruitful than others. Some educational researchers have attempted to determine the nature of the relationships between teacher behavior and student achievement; and, despite tremendous inconsistency across studies and methodological problems that are not inconsiderable, some patterns do emerge. Rosenshine and Furst (1971), for example, reviewed some 60 studies, and provide a list of 11 categories of teacher behavior which appeared with relative consistency to result in significant student gain: clarity, variability, enthusiasm, task-oriented and/or businesslike behaviors, student opportunity to learn criterion material, use of student ideas and general indirectness, criticism, use of structuring comments, types of questions, probing, and level of difficulty of instruction.

Both the Rosenshine and Furst (1971) and the Rosenshine (1971) review on which it was based have been the targets of criticism, much of it justified (cf. Heath and Nielson, 1973; Gall, 1973). The differences across studies in subject populations, content of lessons, definitions and measures of both teacher behavior and student achievement, all undeniably exist and sorely hamper attempts to draw meaningful conclusions regarding causal — or even correlational — relationships between teacher behavior and student achievement. Neither the literature reviewed by Rosenshine nor the studies reported since Rosenshine's reviews could reasonably be used as the basis for making decisions about teacher evaluation procedures or the redesign of teacher education programs; on the other hand, no one except the critics of this research even suggests that any one proposes to do so. Their plea is instead for more research designed to overcome the weaknesses in previous studies; the most they claim for the studies they review is that they suggest the directions which new research might follow.

The major difficulty in interpreting the existing literature lies in the lack of similarity across studies in several important areas. "Student achievement" measures, for example, have included both standardized tests (Stanford Achievement Tests, Wide Range Achievement Tests, Iowa Tests of Basic Skills, and several others) and special tests, some developed specifically for one study (e.g., Fortune, 1967; Wright and Nuthall, 1970). Even when the same tests are used, the results are often analyzed and reported differently; Evertson and Brophy (1973) and Soar (1972) both used the Metropolitan Achievement Tests, but Evertson and Brophy report separate subscale scores, while Soar reports factor scores based on a pooled analysis of the M.A.T. data with scores on other tests.

Similar differences exist in the definition and measurement of teacher behavior. Both the Evertson and Brophy (1973) and the Soar (1972) studies discussed above focused on (among other teacher behavior variables) classroom management and control; but Soar's definition of "control" was based on factor analysis of observational data, Evertson and Brophy used a relatively high-inference rating of an already defined variable. Comparison of these studies thus requires scrutiny of the equivalence of both the definitions and the measurement of "control"; this equivalence cannot be assumed, and is difficult to demonstrate.

Other crucial dimensions also vary from one study to another. The content of lessons ranges as far as the school curriculum; the conditions under which students and teachers interact range from natural, uncontrolled classrooms to highly controlled microteaching situations.

The number of variables studied varies widely; the literature contains both univariate (e.g., Hughes, 1973; Friedman, 1973) and multivariate (e.g., Evertson and Brophy, 1973; Soar, 1972) studies. Both methodologies clearly have their place in the pursuit of knowledge; unfortunately, not all researchers have used these techniques appropriately. Multivariate studies are well suited to the generation of hypotheses in field settings, particularly when the researcher wishes to examine interactions among variables. Univariate studies are particularly useful for hypothesis testing; but such studies must be carefully designed to control for the possibility of interactions between the variable being studies and other variables (sometimes referred to in a revealing way as "nuisance variables").

Still other problems are methodological. Assumptions underlying significance tests are violated (e.g., Evertson and Brophy, 1973), with the result that little trust can be placed on claims of significance. Other investigations omit control groups (e.g., Roberson, 1969), making it logically impossible to attribute differences in student achievement to observed differences in teacher behavior; still others use inappropriate significance tests (for example, Willson, 1973, used multiple t-tests to measure the significance of differences between sets of mean scores, which would be appropriate only if all such comparisons could be shown to be independent, which in this case they were not).

An important point that is almost universally ignored is the amount of correspondence between the student achievement measures used in a study and the educational objectives of the teachers. If, for example, tests sample students' performance on making inferences from their reading, but their teachers are focusing on teaching decoding skills, or perhaps are trying to instill in their students a love of reading, the meaning of correlations between teacher behavior and student achievement becomes at best ambiguous.

It would be counterproductive in this context to delve more deeply into the difficulties in interpreting previous studies. The above-mentioned criticisms have been developed in detail by such writers as Heath and Nielson (1973) and Gall (1973). It is sufficient for our purposes to note that these problems exist, making it necessary for conclusions to be drawn from the literature only with great caution.

The problems which confronted earlier reviewers, and which have been so capably described by their critics, have not gone away. In one sense they have grown worse; for each new addition to the literature brings yet another set of teacher behavior variables, of indeterminate comparability to earlier and concurrent work. A more optimistic interpretation of the situation is possible, however: if several investigators, working in different subject areas with different measures of teacher behavior, have studied variables that are conceptually similar and have found significant correlations between particular categories of teacher behavior variables and student achievement (however measured), these categories of teacher behavior are worthy of further investigation and should be measured in any comprehensive study of classroom processes.

Examined from this point of view, the recent research literature helps us to expand the list provided by Rosenshine and Furst (1971). For example, teachers' use of various reinforcement techniques has been studied in three recent reports. Friedman (1973) examined the relationship between teacher reinforcement of spontaneous student talk and both teacher-initiated student talk and spontaneous student verbal activity within the classroom. Teacher reinforcement of spontaneous student talk was measured on a six-point observational scale from the Observation Schedule and Record (OSCAR) (Medley, 1965) according to a system developed by Medley, Impelliteri and Smith (1969). The categories of the scale are:

(Continued on page 3)

Teacher . . . *(continued)*

1. Supported. The teacher enthusiastically indicates that what the child has said is correct or acceptable.
2. Approved. The teacher clearly indicates that what the child has said is correct or acceptable, but enthusiasm or praise is not given.
3. Acknowledged. The teacher makes some response to what the child has said which does not clearly indicate whether it is correct (acceptable) or incorrect (unacceptable).
4. Not evaluated. The teacher makes no response to a child's verbalizations.
5. Neutrally rejected. The teacher clearly indicates that what the child has said is incorrect or unacceptable, but no disapproval is expressed.
6. Criticized. The teacher clearly indicates with disapproval that what the child has said is incorrect or unacceptable.

Repeated observations were made of interactions between the teacher and four students in 18 classes at each of the following grade levels: first, third, fifth, and seventh. The results showed that teacher reinforcement of spontaneous activities was significantly correlated with the frequency of student-initiated talk both to the teacher and to other students.

With 60 sixth-grade students, Rosenfeld (1972) made successive use of regular classroom reinforcement (e.g., teacher praise or thanks), chart reinforcement (student success displayed on chart in the classroom), and monetary reinforcement for passing specially constructed 25-minute tests on portions of the regular sixth-grade arithmetic curriculum. The number of tests passed under regular classroom reinforcement was compared to the number passed under chart, monetary, and monetary plus chart reinforcement. Lower I.Q. students showed no improvement in any condition, but there was a significant improvement for the total class and for the middle I.Q. students during monetary plus chart reinforcement. High I.Q. students improved under both monetary and monetary plus chart reinforcement.

A critic of these two studies might note that Friedman (1973) measured student behavior but not student achievement, while the reinforcement techniques employed by Rosenfeld (1972) might be somewhat artificial and awkward under normal classroom conditions. Similar findings, however, were also reported by Hughes (1973); in 13 seventh-grade classrooms he experimentally manipulated teacher reactions to students' answers in three specially designed 40-minute science lessons. Lessons were planned in detail and memorized by the teacher. Lesson content was held constant while teacher reactions to student responses were manipulated; student responses were either praised or reacted to neutrally. Student achievement was measured by a specially developed 222-item test. The results showed that positive teacher reactions facilitated student achievement significantly more than minimal (neutral) reactions.

Support for the significance of a dimension of teacher control may be found in two multivariate observational studies. Soar (1972) factor analyzed student achievement in 151 classrooms at the kindergarten, first-, and second-grade levels. The observation instruments used to measure teacher behavior included the Florida Climate and Control System (FLACCS), the Teacher Practices Observation Record (TPOR), and two other observation instruments. Global ratings were also used. Factor analysis of the Florida Climate and Control (FLACCS) identified nine factors: strong control (teacher managing the classroom through coercive control methods and negative affect); pupil free choice vs. no choice (the degree of freedom the pupil has to choose what he wants to do); teacher-pupil supportive behavior (the obverse of strong control — represents teacher management behavior that is gentle and noncoercive); nonverbal gentle control (self-explanatory — e.g., touches, pats); gentle control (gentle verbal control through questions, along with nods, smiles); work without the teacher (students structured into seatwork or a group without the teacher); pupil negative affect (active pupil negative behavior, apparently in the absence of teacher control — e.g., damages property, threatens, teases); teacher attention in a task setting (individual attention from the teacher in task-oriented situations); and teacher positive affect (teacher "bubbly" behavior — outgoing, positive behavior). The results indicated high positive correlations between two pairs of factors (strong control with pupil negative affect, and pupil free choice with teacher positive affect), but factor scores are nevertheless reported separately.

Student achievement measures included the Wide Range Achievement Test, the Metropolitan Achievement Test, and the Stanford Achievement Test; results are reported as factor scores representing "simple-concrete, skill, and complex-abstract" achievement. No summary of the observed relationships between teacher classroom behavior and student gain — or indeed of the study as a whole — will adequately convey the essence of Soar's work; of particular interest in the present context, however, are the following trends:

1. Greater amounts of teacher control, structure, focus and convergence, and lesser amounts of pupil freedom, exploration of ideas, and experimental teaching led to increased student cognitive growth, especially in the skill measures.
2. (It was suggested that) "front and center" teacher behavior is not effective, but rather that structuring of pupil learning activities is related to gain.
3. The amount of structured activities was related to gain as was teacher involvement with pupils in structured activities (adapted from Soar, 1972).

The major pattern that emerges from Soar's findings is the importance of the control and structure dimensions. Learning environments must be structured so as to maximize the opportunity for student growth to occur. This can happen either through active intervention by the teacher (teacher control) or through the student's internalizing a structure which will bring about the same effect. In the early grades, we would expect that this internalizing process has to be facilitated by the teacher.

A somewhat similar pattern of results was reported by Evertson and Brophy (1973). Their research contains methodological flaws which require caution in interpreting their findings; specifically, they studied 31 classrooms, but applied over 1000 measures in all — a procedure which violates the assumptions underlying the significance tests they use. With caution, then, consider their results in comparison with those reported by Soar (1972) and discussed above. Evertson and Brophy's major finding was a significant positive correlation between the teacher's classroom management and control and student gain as measured by the word knowledge, word discrimination, reading, and arithmetic subtests of the Metropolitan Achievement Tests. "The teacher who is well-organized, who monitors the class regularly and nips potentially serious problems in the bud, and who has well established routines for handling everyday procedural matters tends to be more successful in producing learning gains . . . It appears that a key factor in effective teaching is organizing the classroom environment so that there is maximum opportunity to learn."

In the area of teacher questioning behavior, the literature reviewed by Rosenshine and Furst (1971) showed inconsistent results that prohibited any firm conclusions save that multiple-level systems for classifying questions or types of discourse showed more promise than the two-class systems which predominated in the literature. (The two-system types differ in terms of the number and precision of the categories into which they classify questions. Two-class systems classify questions as either lower order or higher order — for example, as asking for information or asking for explanations. Multiple-level systems provide a hierarchy of cognitive levels, allowing finer discriminations among questions; "higher-order" questions, for example, might be subdivided to

(Continued on page 4)

include analysis, synthesis, and transformation questions.) This suggestion received some support from the work of Willson (1973), who studied the level of pupil responses to teacher questions as measured by an interaction analysis system based on Bloom's taxonomy. The study is of particular interest because it employed an experimental manipulation of the teacher's questioning behavior. Observations of both teacher and pupil verbalization levels were made during discussion lessons both before and after the experimental group received training in an interaction analysis technique which focused on their level of questioning behavior. A comparison of the mean levels of teachers' and pupils' verbalizations (as measured by classroom observation) showed that the change in mean levels of thinking in teachers and pupils in the experimental group was significantly higher than that for the control group. Furthermore, there was a direct relationship between the mean levels of thinking of the teachers in both groups and the levels of thinking for the corresponding pupils. The research design used in this study (pre-post, control-experimental) permits two important conclusions to be drawn (with some caution, due to the use of an inappropriate significance test): the training received by teachers in the experimental group systematically altered their level of thinking as reflected in their questioning behavior; and the level of student responses, which presumably reflects the level of their thinking, was significantly affected by the level of teacher questioning. Such a finding has an important implication for a study of the teaching of reading and mathematics; specifically, since both areas have as one objective the raising or development of students' inferential skills, it may be important for teachers to ask higher-level questions in order to develop these skills.

A different effect of teacher questioning behavior — in a sense, a different type of questioning skill — was examined by Fincke (1968). Working with a group of third-grade children, he explored the effect of asking questions to develop purposes for reading on the attainment of higher levels of comprehension. Despite the small size of his sample ($n = 32$), his results reached the .02 level of significance; asking questions to develop purposes for reading resulted in significantly higher levels of comprehension. These findings need to be replicated using a full experimental design; nevertheless, they also indicate a category of teacher behavior that should be observed in future studies of teacher behavior — student achievement relationships.

Gall (1972) reviewed the literature relevant to teachers' use of questions in the teaching of reading. His review supports the conclusions drawn by Rosenshine (1971); teacher questions apparently do affect student achievement, but results are far from consistent from one study to another. Gall's review does document the often reported observation that teachers generally make relatively little use of higher-order questions; instead, the emphasis — in fact, overemphasis — is on information objectives rather than on higher cognitive objectives. One problem in the research literature is that this pattern is too often repeated in the dependent variable measures selected by the researcher. If the measure of student achievement is a multiple-choice, factually oriented test (as was the case with Wright and Nuthall 1970, cited by both Gall and Rosenshine), it is not possible to measure the effects of teachers' higher-order questions, and no conclusions should be drawn from such studies regarding the effectiveness of teacher questioning at various levels.

Overall, the research literature relating teacher behavior to student outcomes might validly be compared to the apocryphal blind men attempting to describe an elephant. Each study examines only a highly limited segment of an extremely complex interactive process; it is hardly surprising that such research has failed to capture the essence of the teaching act.

The literature on the relationships between teacher behavior and student achievement might be examined in terms of a three-dimensional model of teacher behavior. Teacher behavior may be described as related to environmental structuring (Dimension 1), interpersonal relations (Dimension 2), and the goals or objectives of the instructional process (Dimension 3). There is in the existing literature an overwhelming emphasis on the first two dimensions; very little has been reported regarding teacher behaviors that are specifically goal directed (i.e., focused on educational objectives in specific areas, e.g., teacher behaviors involved in teaching decoding skills). For example, of the 11 categories of teacher behavior selected by Rosenshine and Furst (1971) as "most promising," only one — student opportunity to learn criterion material — can reasonably be said to load heavily on Dimension 3. This is one area in which the results are reasonably clear and consistent; correlations between measures of opportunity to learn and student achievement are positive, significant, and consistent. But the general area of teachers' goal-directed behavior has received far less attention than any other.

Two exceptions to this rule have involved evaluative studies such as *Minicourse 18: Teaching Reading as Decoding* (Skailand, 1973; Strickler, 1973). This course, a product of the Far West Laboratory for Educational Research and Development, focused on 26 specific teacher behaviors thought to be involved in teaching reading as decoding. The main field test of this minicourse (Skailand, 1972; Skailand, 1973) showed that Minicourse 18 was effective in changing teacher behavior in these areas. Student achievement data are referred to but not reported; the Far West Lab, however, stated informally (Morris Lai, personal communication, 1973) that the data were inconclusive. Significance levels were not reported, nor was significance claimed, but there was a tendency among first-grade students for the experimental group to show greater gain in decoding skills than did students in the control group. This tendency did not appear at the second- or third-grade levels.

Strickler (1973) also studied the effects of Minicourse 18 on both teacher behavior and student achievement. His findings (in a pre-post, control-experimental design) also showed that Minicourse 18 was effective in shaping teacher behavior; the mean frequency of experimental group teachers' use of 24 of the listed behaviors increased or decreased in the expected direction between pre- and post-course lessons, with the amount of mean change being statistically significant for 16 of the behaviors. When student achievement data were examined, it was found that experimental group students made significantly greater gains in reading achievement, as measured by the Paragraph Meaning and Word Study Skills subtests of the Stanford Achievement Test and the Minicourse 18 Achievement Test, than did students in the control group. It was also reported that less able students made greater gains in reading decoding skills than did more able students, indicating that these teacher behaviors may be most effective with lower ability students.

This review has focused on those studies in the research literature which (1) measured or manipulated teacher behavior, (2) measured student achievement, and (3) studied the relationship between these variables. Emphasis was placed on studies of reading and mathematics instruction at the elementary school level. It has not included studies from related bodies of literature, some of which might be relevant; for example, research into basic reading processes (e.g., Levin and Mitchell, 1969; Calfee, Lindamood and Lindamood, 1973). Also excluded were studies which compared the effectiveness of instructional programs without measuring the behavior of participating teachers (see Corder, 1971, or Chall, 1967, for comprehensive reviews in this area). Although it is probable that a review of the literature in related areas might allow

(Continued on page 5)

Teacher . . . *(continued)*

inferences to be drawn regarding desirable teacher behaviors, the effort required would far exceed the benefits to be gained; it is clearly beyond the scope of this review.

Recommendations

1. Relational studies of teacher behavior and student achievement should include measures of those teacher behavior variables which the research discussed above delineated as clearly worthy of further study. In addition, such research should include measures of Dimension 3 (teacher goal-directed behavior), including but not limited to those behaviors discussed by Skailand (1972) and Strickler (1973).

2. All three dimensions of teacher behavior should be examined together in order to assess interrelationships among them.

3. Observational and analytic systems should be designed so as to allow for the possibility of examining sequential relationships between teacher and student behavior.

4. The content of teacher-student interactions, as well as their quality, should be assessed. For example, it is important to know not only how often teachers reinforced student responses, but also what kinds of student responses were reinforced and what type of teacher reinforcement (e.g., praise, use of student ideas) was employed.

5. Research designs should accommodate all of the variables which affect student achievement. Neither teacher behavior nor student achievement occurs in a vacuum; instead, they occur in a context of complex interaction with other variables, such as organizational climate, teacher and student characteristics, and student peer group behavior.

6. The range of student achievement measures should be expanded to include all levels and types of student gain that are desired by the educational system.

7. The objectives pursued by individual teachers should be assessed. If teachers are focusing their efforts on other objectives than those endorsed by the system, the researcher must know this in order to understand his results.

8. Teacher behavior measures should focus not only on teacher behavior toward the class as a whole, but also on teacher behavior relative to individual students. If teacher behavior is directed only toward some subset of the students in the class, correlations between teacher behavior and mean class achievement are, to say the least, difficult to interpret.

Measures of Teacher Expectations[1]

by

Marlaine Lockheed-Katz

Experimenter bias, operationalized in the classrooms by Rosenthal and Jacobson (1968) as teacher expectancy, has provoked a variety of studies to investigate whether (or how) such expectancy may influence students' performance. Reviews of teacher expectancy literature shed little consistent light on the answer to this question (Finn, 1972; Baker and Crist, 1971; Hidde and Rosenshine, 1973). Rather than concentrate on the findings of these studies, this review will concentrate on how such teacher expectancy has been assessed and manipulated.

Teacher expectancy literature may be divided into four distinct groups: (1) studies in which teachers' expectancies for particular pupils are experimentally manipulated to investigate how such a manipulation will affect student outcomes, such as measured growth in academic achievement or intelligence; (2) studies in which teachers indicate their relative expectancies for the academic achievement of the students in their classroom and this evaluation is correlated with student and/or teacher performances; (3) studies in which teacher expectancies are inferred from student characteristics unrelated to their academic achievement and this inferred expectancy is related to teacher or student performances; and (4) studies in which groups of teachers make evaluations of hypothetical students to inyestigate whether teachers in general hold differing expectations for the performance of certain types of students.

The first group of studies does not utilize a measurement of teacher expectancy. Rather, these studies assume that the experimental manipulation has in fact modified the teachers' expectancies. There are two major forms of experimental manipulation: (1) informing the teacher that the students have been tested and are either "bright" or expected to show academic "growth" or "blooming"; (2) providing the teacher with bogus I.Q. scores.

The second group of studies typically asks the teacher to rank order the pupils in his or her class according to "academic achievement." These rankings are made during the first few months of the academic semester and may be as much a measure of the students' actual performances to date as of the teacher's expectations *per se*.

The third group of studies infer teacher expectancies from observed differences in teacher behavior directed toward different types of pupils. Some characteristics of students, unrelated to academic ability, which may produce differing teacher expectancies, are suggested.

The fourth group of studies measures teacher expectancies as indicated by differential evaluations made of hypothetical students who differ on a variety of characteristics. This strategy has not been reported as an individual measure of teacher expectancies, but rather as a measure of general expectancies held by groups of teachers.

1. Experimental manipulations of teacher expectancies by experimenters

a. *Global evaluations*

In a number of studies related directly to the Rosenthal and Jacobson study, experimenters have told teachers that randomly selected groups of students in their classes were expected to "blossom" academically or were simply "brighter" than their classmates. Dependent variables in

(Continued on page 6)

[1] In preparation for this review, a computer search of the literature was undertaken, using the Lockheed Information Retrieval Service accessing both ERIC and PADAT. Key words were "Expectation," "student teacher relationship," "academic achievement," and "teacher attitudes." A total of 95 articles and dissertation abstracts were retrieved. In addition, three literature reviews on the topic of teacher expectations were located.

Measures . . . *(continued)*

these studies have been both teacher behaviors, student intelligence or achievement growth, or combinations of the two. Results of such studies are inconsistent.

In a study which predates Rosenthal and Jacobson (1968), Flower (1966) randomly assigned two of four lower ability groups of the seventh-grade students higher section " labels. No differences were found in student achievement between the treatment and control groups.

Using experimentally modified files containing reports regarding Head Start students' ability, Beez (1968) found both that teachers behaved differently toward high expectancy students and that the high expectancy students achieved more than the control students.

In Rosenthal and Jacobson's famous study, *Pygmalion in the Classroom*, the researchers advised teachers that randomly selected first-, third-, and fifth-grade pupils had been identified as " potential intellectual bloomers " by their performance on intelligence tests administered to all pupils. First-grade experimental pupils showed greater I.Q. gain than did first-grade controls. This finding was challenged by Elashoff and Snow (1971), a challenge rebutted by Rosenthal and Rubin (1971).

Claiborn (1969) attempted to replicate the Rosenthal and Jacobson study, but expanded the previous design to include observations of teacher behavior. No difference was found in either teacher behavior directed toward " bloomers " or in the I.Q. growth of " bloomers " compared to control subjects within the same classroom. According to Rosenthal and Rubin, however, in an early version of the Claiborn study, the author admits that in two of the six classes in which teacher expectancies were manipulated, the teachers failed to believe the manipulation. In a study of adults in a correctional institution, Krill (1969) told instructor-supervisors that a random 30 percent of subjects tested on the MMPI had shown a " high potential for intellectual gains." After three months, the experimental subjects showed greater I.Q. gains than the control subjects.

Michenbaum, Bowers and Ross (1969) supplied the four teachers of 14 delinquent adolescent girls with a list of " bright " students and had the teachers discuss the students' potentials. Differences between experimental and control groups were found in both teacher behaviors and student academic growth.

Jose and Cody (1971) were unable to establish the modification in teacher expectancy required for a replication of Rosenthal and Jacobson. Eleven of the 18 teachers in their study reported that they had not expected the " academic blooming " of the four experimental students in their class which was " predicted " by the students' scores on a standard achievement test. The researchers found no differences in either teacher behavior or growth in I.Q. between experimental and control students in the same class.

On the other hand, Kester and Letchworth (1972) found differences in teacher behaviors directed toward " bright " students relative to their " average " peers, when teachers were told the identity of the " bright " students.

In a microteaching rather than a naturalistic setting, Rothbart, Dalfess and Barrett (1971) established the expectancies of the teachers by telling them that " Class tests, I.Q. scores, and teacher evaluation of these students indicate that *Student A* and *Student B* have considerably greater academic ability than *Student C* and *Student D* who lack intellectual potential." The researchers found that there were differences in teacher behaviors directed toward high expectancy students and differences in the activity of high expectancy students relative to the low expectancy students.

Of the eight studies reviewed above, which attempted to establish teacher expectancies experimentally by identifying certain students as " bright," " potential intellectual bloomers," having " high potential for intellectual gains," or by providing other global evaluations of the students, two studies found on effect either on teacher behavior or student performance, three studies found effects on both teacher behavior and student performance, one study found effects on student performance but none on teacher behavior, and one study each found both effects and no effects on student performance alone. Rather balanced, if inconclusive findings.

b. *I.Q. scores*

Another commonly used method of establishing teacher expectancies is to provide the teacher with I.Q. scores falsely inflated or deflated for the experimental subjects. This procedure is done both in connection with other information and independently. I.Q. scores in natural classroom settings are modified from 10 points (Pitt, 1956) to 16 points (Maxwell, 1970; Fleming and Anttonen, 1971). Scores may be associated with a label, such as " gifted " or " regular " (Rubovits and Maehr, 1971), combined with grades, achievement test scores, and teacher ratings (Carter, 1970). Teachers may be given the information once or reminded several times (Goldsmith and Fry, 1971).

Differences in teacher behaviors directed toward high expectancy students are reported by Rubovits and Maehr (1971). Differences in student I.Q. gain in favor of high expectancy students are reported by Carter (1970) and Maxwell (1970); no differences are reported by Goldsmith and Fry (1971). Four studies report no differences in achievement gain in favor of high expectancy students (Pitt, 1956; Carter, 1970; Maxwell, 1970; Fleming and Anttonen, 1971).

2. Naturalistic measurements of teacher expectancies in the classroom: rankings

The preceding review of experimental studies, in which teacher expectancies have been manipulated, suggest several bases for measuring teacher expectancies. Teacher may hold expectancies for student performance relative to other students in the class (" bright "), relative to a national norm for ability (high I.Q.) or relative to past performance (" bloomer "). Thus measurements might include the ranking of students in class along some expected achievement dimension, rating students' expected achievement against some absolute norm, or either ranking or rating the students' expected academic growth. In fact, only one of these strategies has been widely used, and that is the teacher ranking of students within the class along a dimension of expected academic achievement. This measurement of teacher expectancy is used by Silberman (1969), Brophy and Good (1970), Mendoza (1971), Cornbleth (1972), Good and Brophy (1972), Jeter and Davis (1973), and Willis (1973). In some studies teachers are requested to rank students in terms of " expected academic achievement," while in other studies teachers are requested to rank students only in terms of " academic achievement." Rating rather than ranking is used in only one study (Davis and Slobodian, 1967), in which teachers were requested to rate students according to their motivation and readiness for reading, " considering the child's ability to pay close attention to explanations, stay with a task and ask relevant questions." The naturalistic studies tend to be concerned with teacher behavior as a major dependent variable. Six of the above eight studies are concerned with differences in teacher behavior directed toward low expectancy vs. high expectancy students. The other two studies sought to identify characteristics of students more likely to be perceived as high or low expectancy students.

(Continued on page 7)

Measures . . . *(continued)*

Silberman (1969), Brophy and Good (1970), Mendoza (1971), Good and Brophy (1972), and Jeter and Davis (1973) all report differences in teacher behavior directed toward high expectancy relative to low expectancy students and differences in the behaviors of the high expectancy students themselves. The consistency of this result, which is at variance with the inconsistency of experimentally manipulated teacher expectancies, suggests that the rankings may be more appropriately considered student measures than teacher expectancy measures. Williams (1972) presents evidence that students' past achievement affects teacher expectations more than teacher expectations affect student achievement. It may be, therefore, that a pattern of student activity, based upon the student's perception of his or her own past achievement, explains the differences in teacher behavior, these studies reveal. None of these studies attempt to control for student achievement while studying the effects of teacher expectancy. Further studies utilizing this form of measuring teacher expectancies should control for student achievement.

3. Inferring teachers' expectancies

A number of studies have been identified in which teacher expectancy is inferred rather than measured. We include these to suggest additional dimensions along which teachers may form expectations for student performance, although the dimensions may be irrelevant to academic achievement.

Rist (1970) conducted an informal study of one kindergarten class of urban black children as they progressed through school. The kindergarten teacher received informal information prior to the school opening regarding the child's family, family income, reputation, older brothers' or sisters' performance in kindergarten, and whether the child had preschool experience. During the first week of school the teacher was able to observe the child's physical appearance, interactional behavior, and language usage. After 8 days of school the teacher seated children permanently, with the middle class children at a table toward the front of the room and the poorer children at the rear. Rist infers that student social characteristics created an expectancy in the teacher for student success which led her to structure the class in such a way that her expectations would be confirmed. Rist's model implies that variance in teacher behavior is an indicator of differential expectancy and that variance in behavior which is accounted for initially by social factors will ultimately produce confirming academic evidence.

Marlaine Lockheed-Katz

Another study in which teacher expectancy was inferred was conducted by Seaver (1971). In a quasi-experimental design, Seaver located 79 first-grade students who were being taught by the same teacher who taught an older sib. Older sibs were judged to be high or low performers in first grade and this " expectancy " rating was applied to the younger sib. Control subjects were children who did not have the same teacher as did their older sibs but whose sib's performance was also judged to be high or low. Seaver inferred that the older sibs' previous performance would set a condition of positive or negative teacher expectancy for the younger sib. The results of the study were as expected; high expectancy students performed better than the controls, while low expectancy students performed less well, holding constant the performance level of the older sib.

4. Measuring the expectancies of groups of teachers

A final group of studies has attempted to measure the generalized expectations held by teachers relative to the performance of certain types of pupils. These studies are consistent with a large body of research documenting that performance expectations are associated with socially evaluated status characteristics, such as race, gender, attractiveness, and so forth. The two studies reported here show findings consistent with previous research, but utilize instruments which measure group rather than individual expectancies.

Clifford and Walter (1973) examined the effect of physical attractiveness on teachers' estimation of the child's I.Q., social relations skills, parents' attitudes toward school, and level of expected academic school completion. A sample of 404 teachers each evaluated one " student file " composed of a student picture and a report card. The report cards were identical for all students; half the students were attractive and half were unattractive. The attractive students were rated higher on all four categories. The researchers conclude that teachers as a group are predisposed to evaluate attractive children more highly than unattractive children.

In a similar vein, Finn (1972) provided a group of 300 teachers each with one of four essay pairs variously attributed to a black or white, male or female fifth-grade pupil with high or low I.Q. and achievement. Teachers were requested to rate the essays on ten 10-point scales. Suburban teachers did not as a group rate the essays differently according to the nature of the hypothetical author. On the other hand, urban teachers rated white female essays significantly lower than white male essays.

Summary

This review has focused on studies which have related teacher expectancy to teacher behavior and/or student performance. It has not included general measures of stereotypes held by teachers; the assumption has been made that expectancies function at an unconscious level and measures of stereotyped beliefs would not be sufficiently powerful to evoke them.

In general, the experimental findings are inconclusive, but suggest that teacher expectancies function more powerfully for younger pupils than for older ones.

In a naturalistic setting teachers apparently tend to behave differently toward high and low expectancy pupils, but this may be as much a result of pupil behavior as teacher expectancy.

Teachers both individually and as a group may hold differing expectancies for students based on their social class, attractiveness, sibs' performance, and sex.

It is clear that studies of teacher behaviors should take into account what performance level teachers expect from the various students in their classes, as well as what performance level the students have achieved relative to each other. The within-class variation of teacher behavior directed toward students who are perceived to be " brighter " or " less bright " than their peers cannot be overlooked.

The California Beginning Teacher Evaluation Study

by Virginia Koehler
National Institute of Education

Background

The competency-based teacher education movement has been the natural outcome of two important developments in education: the demands for accountability, and the feeling that teacher competencies are the standardized behavior units by which teachers should be educated and certified. Thirty-two states are now in various stages of implementing competency-based teacher education programs. But all of these states must answer two difficult questions prior to instituting their programs: what competencies? and, how are the competencies to be evaluated?

In order to answer the first question, planners must take one step back and provide answers to the following two questions: what *student* outcomes are we interested in? and, what teacher competencies produce the desired student outcomes? While answers to the first question are forthcoming, the second — and most important — question in the formula has yet to be answered.

These questions are presently facing the California Commission for Teacher Preparation and Licensing, which was established by the California Legislature in 1970. Three of the Commission's responsibilities are to:

1. . . . formulate performance standards for general teaching and other educational specialties
2. develop and recommend . . . alternative ways in which to demonstrate qualifications for licensing which are adopted to individual differences in candidates . . . while at the same time protecting against incompetence; and
3. develop objective, independent, verifiable standards of measurement and evaluation of teaching competence as it relates to teacher licensing.

The Commission members realized that in order to carry out their mandate, they first had to determine which teacher behaviors affect student learning in reading and mathematics — two outcome targets of their licensing procedure. The Commission, therefore, applied for and received funds to design a study which would begin to identify important teacher behaviors and characteristics which affect student learning. The results of such a study, they felt, would aid the Commission in developing its licensing system. It would also aid teacher training institutions in planning more effective programs.

The Study

The purpose of the Beginning Teacher Evaluation Study is to identify those teaching behaviors which significantly influence student performance in reading and mathematics. To achieve this purpose, the study was divided into three phases. The outcome for the first, the Design Phase, was the proposal for the second, the Pilot Phase. The third will be planned in part on the basis of the experience and results of the second phase.

Phase I: The Design

A skeleton was developed by the Commission staff members in conjunction with Washington personnel in the late fall of 1972. The design was reviewed and revised at a workshop conference by a panel consisting of education research experts, California elementary and secondary educators, teacher training institution representatives, parents, and a representative from the State Board of Education. The agreed-upon design was turned into a proposal for Phase II — the Pilot Year — and was funded. Phase II began in April 1973, with the research activities subcontracted to Educational Testing Service.

Phase II: The Pilot Year

Phase II has two purposes: 1) to develop the procedures for measuring teacher performance and student growth; and 2) to generate from data on these two classes of factors, hypotheses about their relationships. This phase focuses on experienced teachers and their students in the state of California.

(a) The Research Paradigm. Figure I portrays the research paradigm. The dependent variable is student performance; the independent variable is teacher performance. Other variables influencing each of these variables are teacher characteristics and student characteristics. The analysis will assess the relations among these factors.

(Continued on page 9)

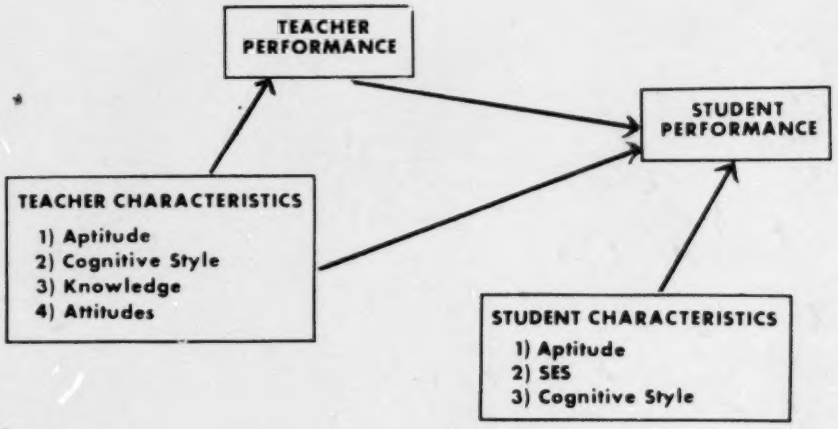

Figure 1. The Research Paradigm

California . . . *(continued)*

(b) Sample of Teachers: Approximately 97 teachers have agreed to participate in this study. Eight school districts were selected on the basis of their urban, suburban, and rural characteristics. Teachers of the second and fifth grades with at least three years of experience were invited to participate in the study. Whenever possible, pairs of teachers from the same school were encouraged to participate.

(c) Criteria of Student Growth: The student performances to be measured are skills and attitudes in reading and mathematics; decoding, comprehension, computation, concepts, application of skills, and attitude toward reading and mathematics. Objectives for the second and fifth grades in each of these categories have been used to construct measures of student performance.

Student performance on these scales is to be measured at two points in time: the fall of 1973 and the spring of 1974. Adjusted gain scores will be used as the measure of pupils' growth with respect to the objectives.

Two kinds of reading tests are being used — a norm-referenced test and a criterion-referenced test, the latter specifically constructed for this study. Measures of decoding skills and attitudes toward reading have also been constructed. An aptitude test is part of the student battery, as well as a measure of students' cognitive style.

(d) Measures of Teacher Characteristics: Measures of four categories of teacher characteristics are being developed: aptitude, knowledge, attitudes, and cognitive style. The measures of teacher aptitude include the measure of verbal aptitude used in the Coleman study, which was a vocabulary test. It also includes measures of memory, classification skills, and inductive and deductive reasoning. The goal of this test development effort is to construct a multifactor test which measures a variety of intellectual aptitudes. Tests of knowledge of decoding processes, of teaching of reading, and of teaching of mathematics have been developed as have measures of teachers' expectancies of performance of children in their classes. Tests of cognitive style and attitudes are being developed.

(e) Measures of Teacher Performance: Observations of classroom performance will be made; several days of observation are currently scheduled. A day of videotaping is also planned. Schedules for two kinds of performances are being constructed: general teaching skills, and those specific to the teaching of reading and mathematics. These measures are being developed by Educational Testing Service according to a model which relates teaching behaviors to specific child achievement goals. A Structured Work Diary is being developed through which teachers will report systematically on their teaching activities. The purpose of the diary is to obtain information on the teaching practices the teachers regularly use.

(f) The Criteria of Teacher Effectiveness: The data will be analyzed in terms of several criteria of effectiveness. One criterion will be the degree to which a teacher's students do better than expected when spring test scores are regressed on fall scores. Discriminant function analysis techniques will be used to identify those performances which characterize those teachers. Another criterion will be to identify the amount of gain a class makes over the year. A third criterion will be the amount of gain of the lowest performing students in the class. Path Analysis will be used to identify the relationships between teacher performances, teacher characteristics, and student characteristics and measures of gain.

Phase III: Hypothesis Testing

Phase III is scheduled to begin in April 1974, and is designed to establish solid empirical relationships between teacher variables and pupil achievement by collecting data from experienced and beginning teachers' classrooms. The variables and hypotheses will be thoroughly tested during this 2-year phase. The sample will be expanded and will include beginning and experienced teachers.

EVALUATION

The Commission has established a Research Advisory Board to advise it on research issues related to the Beginning Teacher Evaluation Study. The Board consists of educational researchers, school administrators, teacher educators, and subject matter specialists.

In addition, the Commission has contracted with an outside firm, Learning Communications Systems, to perform a task audit of their research contractors, Educational Testing Service. The audit will allow the Commission to evaluate the efforts of ETS, and to make management judgments concerning the directions for subsequent phases.

PRODUCTS

The California study is the first statewide project in this country which is attempting to systematically determine teacher behavior variables which affect student achievement prior to establishing a licensing procedure. The information will be utilized by other states, by teacher training institutions in revising their programs. The information will also help to formulate national policy on performance-based evaluation and competency certification.

The results of Phase II will be available in November 1974; and the results of Phase III will be available in November 1976.

Validation of Training Modules In A Competency-Based Program

by Gordon Lawrence and Richard DeNovellis
University of Florida

The Middle School Teaching Competencies Project at the University of Florida has developed 30 training modules for inservice teachers, in a service called Florida Modules on Generic Teaching Competencies. The project, sponsored mainly by the Department of Education at Florida, has field tested all the modules for face validity and has conducted experimental validation studies for seven of them. This report describes the experimental work. Other module developers will be especially interested in two aspects of the report: the use of systematic classroom observation data as the dependent variable in the research, and the successful use of a research design.

Module developers have the problem of validating modules at two levels. First, they want assurance that a module effectively and efficiently assists a trainee in attaining the target competencies during the training cycle. At a second level, they want assurance that the competencies remain in the teacher's on-the-job behavioral repertoire after the training cycle is ended. This research concerns the latter kind of validation of modules.

To date, very little research on modules has been reported. Of 40 reports reviewed, only one by Ricker and Hawkins (2) in-

(Continued on page 10)

Modules . . . *(continued)*

volved an experimental test of significance. They assessed the change in performance of preservice teacher trainees as a result of work with modules.

Research Variables

Seven of the Florida Modules were selected for experimental study because systematic observation was to be used as the means of data gathering, and the seven were thought to have training effects that would be accessible to observation in classrooms at the random times the independent observers would choose to visit.

As dependent variables, seven observation systems were used, one for each of the seven modules. Two of these systems, FLACCS and CASES, had been previously developed ad hoc by the present researchers. FLACCS categories include classroom climate and teacher control variables. CASES consists of 13 categories of student overt behavior.

The independent variables were the seven modules as experimental treatments. See Table 1. Note the FLACCS and CASES were used with all modules. Hypotheses 1 and 2 involve those instruments.

All the modules concern generic teaching competencies — those that pertain to teachers of any subject areas. Because the modules were developed from the same conceptual base, a rationale about middle school education, hypothesis 1 was designed to test the collective effects of all seven modules.

Hypotheses (Null form)

1. When all seven modules are taken together as a single treatment, there will be no significant difference between comparison groups (in the institutional cycle design) as measured by FLACCS and CASES.

2. For each of seven modules separately, there will be no significant differences between the group that used a module and a control group (FI ACCS and CASES observation).

3. For each of seven modules there will be no significant differences between comparison groups as measured by ad hoc observation instruments.

Design: The quasi-experimental design varied according to hypothesis. The institutional cycle design (Campbell and Stanley) was used for hypothesis 1 and for one module covered by hypothesis 3. The diagram shows the design.

Group A	X_1	O_1		
Group B_1 (R)		O_2	X_2	O_3
Group B_2 (R)			X_3	O_4
Group C				O_5 X_4

(O is observation, X is treatment; time interval from X to X was 1 month.)

As the diagram illustrates, the experimental variable (module) is recurrent. The design controls, somewhat more than simple pre-post designs, for sources of external and internal invalidity by making these comparisons: O_2 is less than O_1, O_2 is less than O_3, O_2 is less than O_4, O_5 is less than O_1, and O_5 is less than O_4.

Dunnett's test of planned comparisons was used for hypothesis 2, comparing each mod-

(Continued on page 11)

TABLE 1

INDEPENDENT VARIABLES: MODULES DEPENDENT VARIABLES: OBSERVATION SYSTEMS

		Teacher Behavior		Student Behavior	
Module No.	*Module Concept(s)*	*Ad Hoc Categories*	*General Categories*	*Ad Hoc Categories*	*General Categories*
22–2	Promoting group participation Questioning skills that promote participation	10 categories of classroom structure 5 categories on T-style of questioning	FLACCS		CASES
3–1	Accommodating physical development needs	Adjustment of classroom objects	FLACCS	5 categories of physical activity	CASES
2–3	Supportive teaching strategies	FLACCS	FLACCS	CASES	CASES
5–1	Prompting social-emotional growth	FLACCS	FLACCS	CASES	CASES
17–1	Informal counseling techniques	6 categories of responses to student	FLACCS	7 behavioral responses to counseling technique	CASES
2–2	Growth producing group roles	7 categories of behavior in peer group	FLACCS		CASES
18–1	Helping students clarify values	8 categories of choices open to students	FLACCS		CASES
2–1	Using personal interests of students	20 categories on nature of situation, nature of problem, use of subject matter, evaluation and differentiation	FLACCS	Student questionnaire on interests in re: classroom	CASES

ule group with a control on the basis of FLACCS and CASES data.

For hypothesis 3, one module with a large sample was tested by the institutional cycle design; two modules with small samples were tested by the Wilcoxson sign rank test. For one module the Wilcoxson rank sum test was used. For two others, the ad hoc measurement was served by hypothesis 2. And finally one ad hoc measurement was dropped because observers were unable to find observable situations (module 2–2).

The sample consisted of 135 volunteer teachers in four Florida school systems. One hundred and twenty-seven teachers were in middle schools, grades six through eight; and eight teachers were in an elementary school, grades four through six. Teachers selected modules of interest to them, in the context of a schoolwide individualized inservice program.

Results: Collectively, as a single treatment (Hypothesis 1), the modules influence classroom behavior in a number of dimensions measured by FLACCS and CASES. Module usage was associated with significant changes in these areas: lower aggressive-hostile and lower passive-withdrawn student behavior (CASES); lower verbal and nonverbal negative teacher behavior (teacher verbal and nonverbal *positive* behavior also was significantly lower, presumably because the teachers were less central): teacher less central, making less use of blackboard, attending students more closely, and allowing more student choice. The CASES data comprising "productive behavior" of students changed significantly in two of the five comparisons and that was in the direction opposite to expectations.

The test of hypothesis 2 showed which modules accounted for most of the differences noted above. The relationships were generally predictable. For example, the group that used the module on social-emotional development showed the lowest amount of student hostile-aggressive behavior; and the teachers that used the module on student interests had, as a group, the least teacher-centered classrooms. Almost all the module-using groups differed from the control group by giving the pupils more choice and by making the student more central in the activities of the classroom.

It was possible to test hypothesis 3 for six of the seven modules. Module 2–1, on student interests, had 49 users who were observed. One or more of the comparison groups showed significant differences in four of the five areas of the ad hoc observation categories. (The institutional cycle design proved especially useful because most of the differences showed up on comparisons other than O_2 O_3, the only comparison representing pre- and postobservation on the same persons. In effect, the institutional cycle design demonstrated its power in controlling for sources of invalidity that a simple pre-post design cannot control.)

Module 2–1 usage also involved a pre-post student questionnaire, on the use of student interest in the classroom. Twenty-five classrooms (504 students) were measured. Students perceived increased teacher attention to their interests (.05 level).

The group that used module 3–1, concerning physical development, registered no significant differences from the control group on the ad hoc instrument. One confounding problem was that most of the teachers who chose to use that module were P.E. teachers, and the control group included no P.E. classes.

The ad hoc measure for module 18–1, helping students clarify values, showed a difference beyond the .05 level between treatment and control groups.

The 10 users of module 17–1, informal counseling, were significantly different from a control group on two of six categories of the teacher counseling response instrument.

The group of 12 that used module 2–3, supportive teaching, differed significantly from the control in three categories: teacher verbal positive affect (beyond .01), teacher central (beyond .01), and student productive activity (beyond .05). The latter was opposite the expected direction.

The 14 users of module 5–1, social-emotional development, differed from the control in seven categories: lower pupil aggressive-hostile behavior (.05), less teacher control (.01), more pupil choice (.001), less teacher use of blackboard (.001), less pupil productive activity (.01), and lower frequency of teacher positive verbal and nonverbal behaviors (.01).

Implications

In a naturalistic situation such as the one reported, where treatments and subjects usually cannot be randomly selected, the institutional cycle design is an effective method that offers several comparisons for checking experimental effects. In the case of this study, the design also gives some control over sources of invalidity such as the interaction effects of teacher background and module selection. A disadvantage of the design, however, is its requirement of a fairly large sample size.

In most cases, systematic classroom observation proved well suited to the task of a second-level validation of module effectiveness. However, for some cases, when the target competencies are not regularly used in daily classroom activity, observation periods should be extended for a greater length of time or an alternate method of data collection should be determined. Probably experimental and control teachers could be asked to tell the researcher when target activities would be happening so that an observer could be assigned at that time.

Observation categories need also to be carefully matched to module purposes so as to capture the intended module effects. In this study, the CASES categories did not differentiate student busy work or the *appearance* of constructive work from productive activity.

Finally, this study extended the existing research on modules. Future studies might focus on the further use of systematic classroom observation as a feasible method of validating performance-based training modules. It should be possible to validate some modules with measures of student gains in combination with systematic observation of classroom processes.

REFERENCES

(1) Campbell, Donald T., and Julian C. Stanley. *Experimental and Quasi-Experimental Designs for Research.* Chicago: Rand McNally and Co., 1966.

(2) Ricker, K. and M. Hawkins. "Testing a Science Education Proficiency Module with College Students," *GEM Bulletin,* Vol. 12. Athens: University of Georgia Press, 1969. (ERIC EDO42 695)

(3) Spaulding, R. L. "Coping Analysis Schedule for Educational Settings (CASES)" *in* Anita Simon and E. Gil Boyer, *Mirrors for Behavior II.* Philadelphia. Research for Better Schools, 1970.

(4) Soar, Robert. "Florida Climate and Control Schedule (FLACCS)." Gainesville; Institute for Development of Human Resources, University of Florida, 1971 (mimeographed).

MULTI-STATE CONSORTIUM ON PERFORMANCE-BASED TEACHER EDUCATION

STATE	*REPRESENTATIVE*
Washington	Edwin L. Lyle
Oregon	James Wallace
Utah	Roger C. Mouritsen
Arizona	John Potts
Texas	Tom T. Walker
Minnesota	Patricia J. Goralski
Vermont	Robert Vail
New York	Vincent C. Gazzetta
Florida	Nell Kamnwischer
New Jersey	Ward Sinclair

Administering State: New York

Project Director
Theodore E. Andrews
New York State Education Department
Albany, New York 12210

Stuart E. Dean
United States
Office of Education Representative

Syracuse University Coordinator
James Collins

Editor
Theodore E. Andrews

Funded
by
United States Office of Education
Title V, Section 505, Grant
and
Syracuse University Projects
Supported
by
Teacher Corps
and
National Center For the Improvement
of Educational Systems